ARTICLE IN PRESS

Journal of Petroleum Science and Engineering xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: www.elsevier.com/locate/petrol

A hybrid particle swarm optimization and support vector regression model for modelling permeability prediction of hydrocarbon reservoir

Kabiru. O. Akande[a,*], Taoreed. O. Owolabi[b,e], Sunday. O. Olatunji[c], AbdulAzeez AbdulRaheem[d]

[a] Institute for Digital Communications, School of Engineering, University of Edinburgh, United Kingdom
[b] Physics Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
[c] Computer Science Department, University of Dammam, Dammam, Saudi Arabia
[d] Petroleum Engineering Department, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia
[e] Physics and Electronics Department, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria

ARTICLE INFO

Keywords:
Particle swarm optimization (PSO)
Support vector regression (SVR)
Hydrocarbon reservoir
Permeability prediction
Optimization techniques

ABSTRACT

The significance of accurate permeability prediction cannot be over-emphasized in oil and gas reservoir characterization. Support vector machine regression (SVR), a computational intelligence technique, has been very successful in the estimation of permeability and has been widely deployed due to its unique features. However, careful selection of SVR hyper-parameters is highly essential to its optimum performance and this task is traditionally done using trial and error approach (TE-SVR) which takes a lot of time and do not guarantee optimal selection of the hyper-parameters. In this work, the performance of particle swarm optimization (PSO) technique, a heuristic optimization technique, is investigated for the optimal selection of SVR hyper-parameters for the first time in modelling and characterization of hydrocarbon reservoir. The technique is capable of automatic selection of the optimum combination of SVR hyper-parameters resulting in higher predictive accuracy and generalization ability of the developed model. The resulting PSO-SVR model is compared to SVR models whose parameters are obtained through random search (RAND-SVR) and trial and error approach (TE-SVR). The comparison is done using real-life industrial datasets obtained during petroleum exploration from four distinct oil wells located in a Middle Eastern oil and gas field. Simulation results indicate that the PSO-SVR model outperforms all the other models. Error reduction of 15.1%, 26.15%, 12.32% and 7.1% are recorded for PSO-SVR model compared to ordinary SVR (TE-SVR) in well-A, well-B, well-C and well-D, respectively. Also, reduction of 12.8%, 23.97%, 2.51% and 0.11 are recorded when PSO-SVR and RAND-SVR results are compared in the respective wells. Furthermore, the results show the potential of the application of heuristics algorithms, such as PSO, in the optimization of computational intelligence techniques employed in hydrocarbon reservoir characterizations. Therefore, PSO technique is proposed for the optimization of SVR hyper-parameters in permeability prediction and reservoir characterization based on its superior performance over the commonly employed optimization techniques.

1. Introduction

Permeability is defined as the ease of movement of oil and gas through a porous rock (Olatunji et al., 2014). It is a very important property in reservoir characterization and its accurate prediction is essential to a successful oil and gas exploration. Several decisions regarding the overall management of oil and gas reservoir are made based on the knowledge of permeability. Information such as the scale of the oil and gas present in the reservoir, the amount of recoverable oil, flow rate of the medium, estimate of future exploration and the various exploration equipment and techniques to be employed during the drilling process are supplied based on accurate prediction of permeability (Akande et al., 2015; Tusiani and Shearer, 2007).

It is not sufficient to have oil or gas in the reservoir or formation, the so called ‘oil in place’. Rather, what is paramount is for these hydrocarbons to be able to flow from the formation to the well bore so as to be recoverable at the surface. Permeability, defined as the ease with which fluids flow through the rock, determines this flow rate. Hence, permeability determines the recoverable reserves (amount of recoverable hydrocarbons) from the reservoir volume (oil in place). This makes permeability one of the most important flow characterizations of oil and gas reservoir whose accurate determination is very vital

* Corresponding author.
E-mail addresses: koakande@gmail.com (K.O. Akande), owolabitaoreedolakunle@gmail.com (T.O. Owolabi), oluolatunji.aadam@gmail.com (S.O. Olatunji), aazeez@kfupm.edu.sa (A. AbdulRaheem).

http://dx.doi.org/10.1016/j.petrol.2016.11.033
Received 17 May 2016; Received in revised form 18 October 2016; Accepted 24 November 2016
Available online xxxx
0920-4105/

Nomenclature

C	Regularization Factor
CIT	Computational Intelligence Techniques
PSO	Particle Swarm Optimization
RAND-SVR	Random Search SVR
RBF	Radial Basis Function
RMSE	Root Mean Square Error
SVM	Support Vector Machines
SVR	Support Vector Regression
TE-SVR	Trial and Error SVR
ε	Epsilon

to a successful resolution of many fundamental issues encountered during oil and gas exploration (Akande et al., 2015; Olatunji et al., 2011b).

There are basically three major approaches to permeability estimation which include empirical, statistical and computational intelligence method. The standard method is the empirical approach which involves direct measurement of permeability from core samples obtained from field exploration (Olatunji and El-sebakhy, 2008). This method is very costly and time-consuming especially in the event of large samples which are fairly common during oil and gas exploration. An alternative is the statistical method which involves the use of regression analysis to predict and estimate permeability using several descriptors. This method leads to the development of a linear model which can be used to generalize to unseen data. However, the relationship between permeability and other well logging data used as descriptors is highly non-linear which results in poor performance of the linear model developed from regression analysis (Olatunji et al., 2011a). Hence, the use of computational intelligence techniques (CIT) whose performance surpassed those of the various regression analyses commonly employed becomes paramount.

The special nature of CIT enabled them to adequately learn the underlying non-linear relationship between permeability and the petrophysical parameters used in its prediction. Among the computational intelligence techniques used for permeability prediction, SVR perform excellently well due to its many unique features such as its sound mathematical foundation, non-convergence to local minima and minimization of generalized error bound which ultimately leads to accurate generalization and predictive ability (Akande et al., 2015). SVR is the regression version of support vector machines (SVM) proposed by Vapnik (Vapnik, 1995). SVM is typically employed in classification problems but was extended to the regression case (SVR) due to the introduction of ε-insensitive loss function. The ε-insensitive loss function determines the error-free margin of the model and allow the performance of linear regression in high-dimensional feature space. SVR has been applied to a wide range of real life problems and has performed excellently in cases such as prediction of superconducting temperature, estimation of surface energies, forecasting of stock prices, medical diagnosis and so on (Adewumi et al., 2016; Owolabi et al., 2016a, 2016b; Akande et al., 2016; Olatunji, Elshafei et al., 2011). The optimum performance of SVR depends greatly on the combination of its hyper-parameters which are usually user-defined (Vladimir Cherkassky, 2004).

The most widely employed approaches in choosing SVR hyper-parameters are grid search, random search and trial and error approach (Bergstra and Bengio, 2012; Larochelle et al., 2007; LeCun et al., 1998). Grid search suffers from the curse of dimensionality because the number of parameter combinations grows exponentially with the number of hyper-parameters while trial and error approach present difficulty in result reproduction which is an important component of any research activity (Bellman, 1962). Random search has been shown to match the performance of grid search and outperforms it in many cases with less computational time (Bergstra and Bengio, 2012). The SVR model resulting from random search approach is henceforth referred to as (RAND-SVR) while the model obtained using trial and error approach (manual search) is termed (TE-SVR). TE-SVR is prevalent as the state of the art despite many alternatives due primarily to the fact that it gives quick insight into the problem at hand and requires no technical barrier or overhead (Bergstra and Bengio, 2012).

In this work, the use of Particle swarm optimization (PSO), a global optimization algorithm, in the optimum selection of SVR hyper-parameters is investigated and the developed model is applied to the prediction of permeability in hydrocarbon reservoir characterization. The performance of the resulting model (PSO-SVR) is compared to those of RAND-SVR and TE-SVR models. PSO is an evolutionary optimization technique which combines the principles of social psychology in socio-cognition human agents and evolutionary computations (Eberhart and Shi, 2001; Shi and Eberhart 1998a, 1998b; Kennedy and Eberhart, 1995; Eberhart and Kennedy, 1995). It resulted from hybrid ideas of combining social behavior of organisms such as fish schooling and bird flocking with biological behavior embedded in evolutionary algorithms (Abido, 2002). It offers good blend and balance between global exploration and local exploitation providing solution to the popular exploration-exploitation trade-off problem which results in a flexible and well-balanced algorithm for optimization of parameters. It can be used to solve multi-modal, non-differentiable and non-linear problems with the certainty that the global minimum will be reached since it is a stochastic search of the problem global space which iteratively locate the global minimum of the cost function (Abido, 2002).

It can be stated, after detailed search of the literature and as far as we know, that this is the first time PSO is been deployed to optimize the hyper-parameters of SVR for modelling permeability prediction of hydrocarbon reservoir. The specific contribution presented in this research work include: 1) a detailed investigation of PSO technique in optimization of SVR hyperparameters for hydrocarbon reservoir characterization; 2) a detailed comparison of the developed PSO-SVR, RAND-SVR and ordinary SVR (TE-SVR) models; 3) presentation of important observations as deduced from the research results.

The rest of the paper is organized as follows: Section 2 details the mathematical formulation of SVR and PSO. Section 3 describes the methodology and procedures underlining the results of this paper. Section 4 presents results and discussions while Section 5 states the conclusion and recommendations.recommendation

2. Mathematical descriptions of SVR and PSO techniques

2.1. Support Vector Regression

Support Vector Regression (SVR) is the algorithm derived from SVM by the introduction of ε-insensitive loss function which enables SVR to solve both linear and non-linear regression problems in high-dimensional feature space (Smola and Sch, 2003; Drucker et al., 1996). The input x is first mapped into an l-dimensional feature space using some non-linear function and then linear function is constructed in this feature space (Cherkassky and Ma, 2004). For linear regression problem, the training data is represented as (x_i, y_i), (i=1..., m) where x is a l-dimensional input such that $x \in R^l$ and $\in R$. SVR linear regression model can be represented as (Smola and Sch, 2003):

$$f(x) = \langle \omega, x \rangle + b,\ w,\ x \in R^l, b \in R \quad (1)$$

Where $f(x)$ is the output of linear SVR and $\langle w, x \rangle$ represent dot product between ω and x. SVR minimizes the empirical risk on training data using ε-insensitive loss function proposed by Vapnik and defined in Eq. (2) as follows: